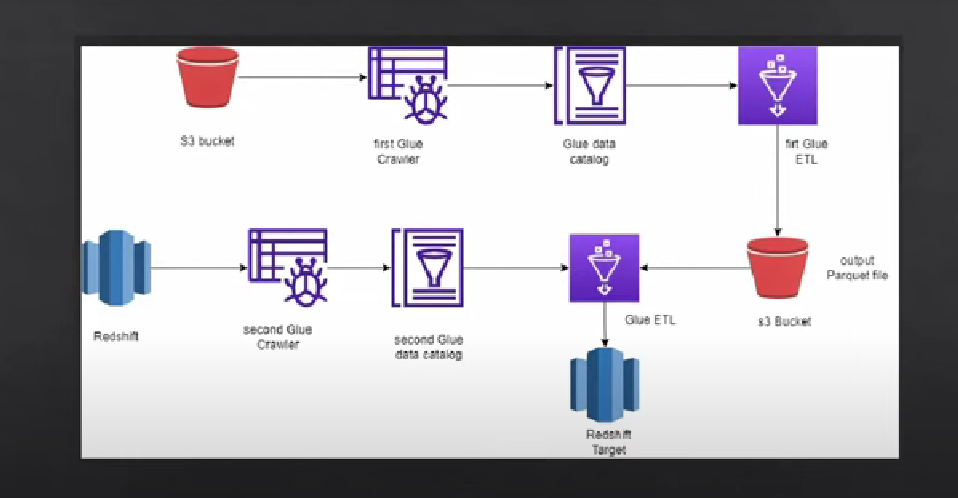
**Glue ETL Project**

1. Architecture Diagram:



Description:

As above flowchart read the customer data from S3 bucket and load the cleaned and transformed data into S3 bucket with optimized parquet File. Load the data parquet file to Redshift data warehouse for analytics.

Services Requirement:

1. S3 Bucket
2. Glue Job
3. Redshift Cluster
4. Crawler
5. VPC
6. IAM roles – S3Fullaccess , RedshiftFullaccess
7. CloudWatch
8. Spark and python library(logger)